

About 3ie

The International Initiative for Impact Evaluation (3ie) promotes evidence-informed equitable, inclusive and sustainable development. We support the generation and effective use of high-quality evidence to inform decision-making and improve the lives of people living in poverty in low- and middle-income countries. We provide guidance and support to produce, synthesise and quality assure evidence of what works, for whom, how, why and at what cost.

3ie working papers

These papers cover a range of content. They may focus on current issues, debates and enduring challenges facing development policymakers, program managers, practitioners, and the impact evaluation and systematic review communities. Policy-relevant papers in this series synthesise or draw on relevant findings from mixed-method impact evaluations and systematic reviews funded by 3ie, as well as other rigorous evidence to offer new analyses, findings, insights and recommendations. Papers focusing on methods and technical guides also draw on similar sources to help advance understanding, design, and use of rigorous and appropriate evaluations and reviews. 3ie also uses this series to publish lessons learned from 3ie grant-making.

About this working paper

This paper, *Integrating impact evaluation and implementation research to accelerate evidence-informed action*, responds to a growing interest in implementation science and faster, less expensive evaluations that provide more timely and relevant information than large cluster randomized controlled trials. Implementation science was originally defined in the health sector but is now being applied in other sectors to address implementation-related questions.

The content of this paper is the sole responsibility of the authors and does not represent the opinions of 3ie, its donors or its board of commissioners. Any errors and omissions are also the sole responsibility of the authors. Please direct any comments or queries to the corresponding author, Anna C Heard, at ahheard@3ieimpact.org.

3ie receives funding for the working paper series from 3ie's donors, which include UK aid, the Bill & Melinda Gates Foundation, and the William and Flora Hewlett Foundation. A complete listing of 3ie's donors is on the [3ie website](#).

Suggested citation: Rutenberg, N and Heard, AC, 2018. *Integrating impact evaluation and implementation research to accelerate evidence-informed action*, 3ie Working Paper 32. New Delhi: International Initiative for Impact Evaluation (3ie). Available at: <https://doi.org/10.23846/WP0032>

3ie Working Paper Series executive editors: Beryl Leach and Emmanuel Jimenez

Production manager: Bawa Sahibjeet Singh

Assistant production manager: Akarsh Gupta

Cover design: John F McGill and Akarsh Gupta

Cover photo: Josemar Franco / Shutterstock. The photograph has been used for illustrative purposes only.

© International Initiative for Impact Evaluation (3ie), 2018

Integrating impact evaluation and implementation research to accelerate evidence-informed action

Naomi Rutenberg
Independent researcher

Anna C Heard
International Initiative for Impact Evaluation (3ie)

Working Paper Series 32

October 2018



Summary

Evidence-informed action requires information about what works, for whom, why and at what cost. Impact evaluations answer questions about efficacy and effectiveness of interventions and programs. They account for factors that may affect an intervention's effect, and they are the only way to attribute the measured effect to the evaluated program or intervention. Implementation research mobilizes theories, concepts and methods to better understand what, why and how interventions work. It studies how the intervention is, and should be, implemented in different contexts, and how to disseminate and scale up the intervention. Implementation research was initially used to study how to introduce and scale up evidence-based practices to improve the quality and effectiveness of healthcare. It has evolved to include the study of myriad factors affecting the process of implementation – bringing an intention into effect.

In this paper, we argue that impact evaluations and implementation research produce different types of equally necessary information for evidence-informed action, and that incorporating research questions and study designs from impact evaluation and implementation research will improve evidence for both types of studies and strengthen program evaluation. Adding more explicit assessments of implementation outcomes along the causal pathway, more rigorous methods to do so, and specifics of how and how well interventions are implemented will improve the usefulness and generalizability of both types of research. We believe prospectively embedding evaluation and measurement in the implementation process is a more robust approach – and probably more efficient, although not quicker – for producing the desired information for evidence-informed action. We focus on research in health, as this is the field from which implementation science, and thus implementation research, emerged. However, the same argument can be made for improving evidence in other sectors, since most programs are at least partially dependent upon how well they are implemented.

We propose an approach to integrating impact evaluation and implementation research that incorporates the explicit recognition and rigorous measurement and assessment of mediators and outcomes from implementation research into the theory of change used for impact evaluation. This framework adopts several elements more commonly used in implementation research, including process evaluation of implementation moderators, a theory of change about the behavior of individuals and organizations implementing the intervention, measurement and reporting of adherence to content and dose, and standardized reporting of the intervention sufficient for replication. Greater rigor in measuring implementation is adopted from impact evaluation. Evaluation of both implementation and beneficiary outcomes can use experimental or quasi-experimental designs.

We apply this framework to three published program evaluations to describe the use of and missed opportunities for integration. We examine three case studies: two impact evaluations from an International Initiative for Impact Evaluation program on the integration of HIV services and one case from the literature, in which researchers and implementers integrated implementation research and impact evaluation to build the case for an innovative approach to long-term care for people living with HIV. We highlight how the integration of these elements contributed or was missed. We conclude with a set of practical recommendations for integration.

Acknowledgments

We would like to thank Sara Pacqué-Margolis and Annette N. Brown for their thoughtful comments on earlier drafts of this working paper and review comments provided by Beryl Leach, Emmanuel Jimenez and Marie Gaarder, which helped improve the final version.

Contents

Summary	i
Acknowledgments	ii
Abbreviations and acronyms	iv
1. Introduction	1
2. Definitions	1
3. Integrating impact evaluation and implementation research	5
4. Case studies	7
4.1 Case study 1: evaluating the impact of community health worker integration into prevention of mother-to-child transmission of HIV services in Tanzania	8
4.2 Case study 2: improving ART adherence at reproductive and child health clinics integrating Option B+ in Tanzania	10
4.3 Case study 3: piloting, evaluating and scaling up community ART support groups in Mozambique	12
5. Conclusions	14
6. Recommendations	15
Appendix: 3ie thematic grant window for integration of HIV services	17
References	18

Abbreviations and acronyms

ART	Antiretroviral therapy
CAG	Community ART group
CHW	Community health worker
PMTCT	Prevention of mother-to-child transmission
StaRI	Standards for reporting implementation studies
TIDieR	Template for intervention description and replication

1. Introduction

Evidence-informed action requires information about what works, for whom, why and at what cost. Before adopting a new drug, technology, service delivery approach or policy, a decision maker should ask three types of questions. First, is the treatment, or innovation, efficacious? Are there well-controlled studies that show the innovation works – that it produces the predicted outcome? Second, is the intervention effective? Will it work in the real world, outside of a controlled environment and for a heterogeneous population? Will it work if program participants (for example, patients) do not show up for all the required visits, if the trained provider transfers to another facility or if implementers interpret the intervention differently in practice? Third, what is the most efficient approach to offering the innovation – that is, attaining maximum productivity with the fewest resources? Can the intervention benefits be increased and costs decreased?

Different methods can be used to answer these questions. In this paper, we argue that incorporating designs and elements from impact evaluation and implementation research will improve evidence for both types of studies and strengthen program evaluation. We focus on research in health because the term *implementation research* (or *implementation science*) has its roots in health, although the same argument can be made for improving evidence in other sectors, since most programs are at least partially dependent upon how they are implemented.

2. Definitions

In a strict sense, impact evaluations answer questions of efficacy and effectiveness of interventions. The terms *impact evaluation* and *randomized evaluation* have often been used in development economics (White 2009; Easterly 2009). In health, the term *pragmatic clinical trial* was used to describe less controlled evaluations in the real world (Patsopoulos 2011; Hemming et al. 2015). Impact evaluations in health assess performance under real-world conditions and account for patient, provider and system factors that may moderate an intervention's effect (Singal et al. 2014). To do this, evaluation designs must take account of selection bias by identifying a comparison group, using either an experimental or a quasi-experimental approach (White 2009; Gertler et al. 2011). We define impact evaluations as studies that measure changes in outputs, outcomes and impacts that occur as a direct result of the intervention. Impact evaluations provide quantitative estimates of changes attributable to the intervention.

Implementation research mobilizes theories, concepts and methods to better understand why and how interventions work in practice (Ridde 2016), usually with the aim of improving or scaling them. At one time, implementation research was heavily focused on understanding why evidence-based practices were underutilized and developing generalizable knowledge to facilitate their utilization. For example, Eccles and Mittman (2006 p.1), in introducing the journal *Implementation Science*, define implementation research in this way:

The scientific study of methods to promote the systematic uptake of research findings and other evidence-based practices into routine practice, and, hence, to improve the quality and effectiveness of health services.

The definition of implementation research has evolved to embrace all aspects of implementation, including the implementation processes and the factors affecting them, as well as how to introduce potential solutions into a health system to promote their large-scale use and sustainability. Implementation research can include formative research, process evaluation, assessment of the fidelity of implementation and quality improvement research, as well as cost-effectiveness analysis or the comparison of two delivery methods. In this paper, we use this expanded definition proposed by Peters and colleagues (2013 p.1):

Implementation research is the scientific inquiry into questions concerning implementation—the act of carrying an intention into effect, which in health research can be policies, programs, or individual practices (collectively called interventions).

Implementation research focuses on organizational behavior – specifically, the uptake of interventions and the adoption of new approaches to delivering care by providers and health systems. It contributes to effectiveness and efficiency studies by identifying optimal implementation options for a given setting.

In health efficacy trials, impact evaluation and implementation research are sometimes juxtaposed on a spectrum (for example, Peters et al. 2013). At one end of this spectrum, we find tightly controlled experimental designs establishing the efficacy of interventions. A common example is a clinical trial that takes a new product or idea with demonstrated efficacy in a lab setting (for example, insecticide-treated bed nets) and tests its effect on people in a highly controlled setting (does it kill mosquitos and prevent them from getting inside the net?).

In the middle of the spectrum are impact evaluations conducted in the real world, exploring questions about the effectiveness of an intervention, on average, or how it might vary in different settings. Impact evaluations control some conditions to minimize selection bias, but in the context of ongoing service delivery. For example, an impact evaluation could assess whether offering insecticide-treated bed nets reduced the number or percentage of children diagnosed with malaria by comparing those provided with nets and a similar group not provided with nets (a counterfactual).

On the other end of the spectrum is implementation research, which informs how to increase the intervention's efficiency – reducing costs or increasing the outputs and benefits – and/or how implementation should be adapted in different settings, as well as how to disseminate and scale up the intervention. Frequently, these are observational studies with nonexperimental designs. For example, implementation research could assess whether, after offering bed nets, a mass media campaign about the importance of bed nets increases use of nets, comparing use before and after. Or, implementation research could assess whether electronic medical records reduce the amount of time needed to fill out administrative forms.

Fundamentally, the difference between impact evaluation and implementation research is in the research questions – efficacy, effectiveness, adoption of, and fidelity to the intervention and efficiency – and not the study methods. However, the two terms are more generally used not only for certain types of research questions, but also for certain

types of study designs. Many of the typical features are listed in Table 1. Although the different features can be, and have been, combined, in this paper we argue for integrating these features more routinely and explicitly to improve evidence-informed decision-making.

Table 1: Typical features of basic impact evaluation and implementation research in the health sector

	Impact evaluation	Implementation research
Research aims	Assess effectiveness of program or intervention for the intended beneficiary	Assess effectiveness of the implementation of the intervention and fidelity of the delivery of the intervention
Research questions	Are there changes in outcomes that occur as a direct result of the program or intervention? <i>For example, does a community health worker program increase retention in prevention of mother-to-child transmission (PMTCT) services?</i>	What are successful approaches to incorporating a program or practice (an intervention) at the community, agency or practitioner level? What are the factors that contribute to adoption and integration of evidence-based interventions? <i>For example, does providing training and weekly mentoring to community health workers improve their knowledge and practice to deliver the expected package of community-based services?</i>
Study designs	<ul style="list-style-type: none"> - Comparison of effect of intervention to counterfactual (what would have happened in the absence of the intervention): <ul style="list-style-type: none"> o Randomized controlled trials (also known as randomized evaluation) in “real world” setting o Quasi-experimental study designs to create a statistical counterfactual, such as difference-in-difference, regression discontinuity, use of instrumental variables or propensity score matching 	<ul style="list-style-type: none"> - Cross-sectional study (for example, comparing two methods of delivery or comparing services delivered in two areas at the end of the intervention) - Pre- and post designs - Case studies - Focus groups and in-depth interviews - Monitoring of routine data

	Impact evaluation	Implementation research
Units of analysis	<ul style="list-style-type: none"> - Beneficiary (outcome) - Facility of catchment area using aggregated beneficiary data at facility level 	<ul style="list-style-type: none"> - Beneficiary (acceptability) - Provider - Facility - System
Comparison conditions	<ul style="list-style-type: none"> - Standard of care 	<ul style="list-style-type: none"> - Implementation as usual or a competing implementation strategy
Sampling frame	<ul style="list-style-type: none"> - Individuals or service delivery units (for example, clinic or catchment area) - Some inclusion/exclusion criteria (for example, non-naïve patients or minimum client load) 	Providers, health facilities
Primary study measures	<ul style="list-style-type: none"> - Adherence to and retention in care - Changes in knowledge and behavior - Clinical/patient/health outcomes 	<ul style="list-style-type: none"> - Facilitation strategy - Monitoring and feedback - Adaptation - Relevance - Changes in: <ul style="list-style-type: none"> ○ Professional behavior ○ Organizational structures ○ Relationships ○ Fidelity (content, frequency, dose)
Challenges	<ul style="list-style-type: none"> - Seen as expensive - Randomization sometimes is not possible - Difficult to make appropriate matched intervention and control sites - May require a lengthy time frame to see results 	<ul style="list-style-type: none"> - Assumes improvements in fidelity will result in improvements in outcomes - Selection bias – results not necessarily attributable - Administrative data, commonly used for this type of assessment, is often of low quality

Incorporating designs and elements from both types of research in program evaluation strengthens the evaluation in at least three ways. First, the attention to internal validity and controlled settings in impact evaluation can ignore the realities of context and produce results that are internally valid but externally irrelevant. Implemented in a different setting, the intervention may not work as well; for example, patients with a different set of characteristics could be less likely to adhere to the treatment or providers might not have the knowledge or tools to provide the intervention as intended. If interventions are to be of greater value, we need to know under which contexts they are most effective and how their effects, or the key components for effectiveness, change in different contexts (Geng et al. 2017).

Second, when impact evaluations are conducted in the “real world” of operational service settings, the researchers cannot control many elements of the intervention. Components may be misunderstood, modified or ignored by providers or the health facility as they integrate them into practice. In the absence of measurement of the fidelity of implementation, this can result in what is often called a Type III error – evaluating a program that is described but not implemented. In such cases, it cannot be known if the lack of an intervention effect is due to deficiencies in the hypothesized theory of change or weak implementation of the intervention.

Third, implementation research studies often use less rigorous methods with nonexperimental designs, which does not allow attribution. This may result in thinking a program is improving when, in fact, the general time trend is causing more people to use it (for example, offering more family planning options and finding that more people are using birth control after one year). And, when implementation research is carried out without evidence from an impact evaluation of the effectiveness of the intervention, the research may suggest ways to improve the quality or quantity of the intervention before it is known whether the intervention makes a difference in outcomes.

In this paper, we propose an approach to integrating impact evaluation and implementation research. We then apply the approach to three published program evaluations to identify where impact evaluation and implementation research were combined and where opportunities were missed. The final section of the paper provides recommendations for combining impact evaluation and implementation research.

3. Integrating impact evaluation and implementation research

Integration of impact evaluation and implementation research is not new. In an editorial in the *Journal of Acquired Immune Deficiency Syndromes*, Padian and colleagues (2011) write that impact evaluation can be used on an ongoing basis to assess intermediary outcomes to ensure a program is on track, or to assess the comparative efficiencies and cost-effectiveness of different programs.

The literature on impact evaluation and implementation research for health programs and interventions presents a variety of approaches for juxtaposing and combining the two types of research. One approach is an effectiveness–implementation hybrid design that gives prominence either to impact evaluation or implementation research questions with findings on a spectrum from “implementation not relevant” to “implementation as primary focus” (Curran et al. 2012; Peters et al. 2013). Another approach incorporates impact evaluations designs – in particular experimental trial designs – into implementation research to enhance its rigor (Patsopoulos 2011; Hemming et al. 2015). A third approach applies a greater use of implementation research elements in impact evaluations to better capture the heterogeneity of the real world in evaluations and bolster the relevance of evaluation findings for diverse decision makers (Habicht et al. 1999; Kalibala et al. 2016; Parry and Power 2016; Geng et al. 2017).

The approach to integration described here proposes embedding research on the determinants of implementation outcomes in evaluations of beneficiary outcomes. Implementation research helps understand how to achieve fidelity of implementation to the intended intervention dose and content. Adding an understanding and measurement

of the fidelity of implementation to impact evaluations prevents potentially false conclusions about an intervention's effectiveness (Type III errors). The measurement, elucidation and reporting of fidelity gives researchers confidence in attributing outcomes to the intervention and about comparability when synthesizing findings. Research and reporting on the fidelity of implementation also provides practitioners with critical information for replicating the intervention properly (Mihalic 2004; Carroll et al. 2007).

Essential implementation outcomes include one or more of the following (Fixsen et al. 2005):

- Changes in professional behavior (knowledge and skills of practitioners and other key staff members within an organization or system);
- Formal and informal changes in organizational structures and cultures (values, philosophies, ethics, policies, procedures, decision-making) to routinely bring about and support changes in professional behavior; and/or
- Changes in relationships to consumers and stakeholders (location and nature of engagement, inclusion, satisfaction).

A large array of contextual factors influences these implementation outcomes, including facilitation strategies, monitoring and feedback loops, the scope for adaptation, and relevance to providers and beneficiaries (Carroll et al. 2007). Facilitation strategies, including guidelines, manuals and training, ensure everyone is receiving the same training and support, with the aim of standardizing the delivery of the intervention. Implementation approaches that provide monitoring and feedback to key stakeholders to encourage learning and adaptation by implementing agencies and beneficiary groups are particularly suitable for working in complex systems (Peters et al. 2013). If providers and beneficiaries do not see the relevance of an intervention (if it is seen as a low priority), or if providers are uncomfortable with the content or perceive a lack of resources (for example, equipment, guidelines, budget or time), then motivation to deliver the intervention will be low, fidelity will suffer and few will access the service (Kalibala et al. 2016).

Fidelity can be assessed in many ways, including structured observations, implementer self-reports and interviews, participant interviews, mystery clients and routine monitoring of data. Quantitative and qualitative analysis can identify mediators associated with fidelity. Evaluators often use process evaluation as a check that what was supposed to be delivered is being, or was, delivered. However, program evaluators frequently miss an opportunity by failing to take advantage of the experimental designs used in impact evaluations to assess and understand the implementation process. Even in the intervention arm of a study, different implementation approaches can be tested in a randomized design¹ (for example, testing the added value of text messages to reinforce provider training among a randomly selected subset of all providers trained).

The theory of change in impact evaluations often incorporates assumptions about the fidelity of the intervention; that is, assuming a certain content, frequency and dose of the intervention being studied. Based on the expected implementation, the theory of change describes the pathways by which the intervention will create changes in beneficiaries'

¹ In health, experimental designs are called randomized controlled trials. For this paper, we use the term more often used in economics – *randomized evaluation* or simply *randomization*.

knowledge, attitudes, behaviors and well-being. However, these assumptions about fidelity are often not explicit and are, in fact, hypotheses. The program evaluator is hypothesizing, for example, that if the provider participates in a training, then that provider will have the knowledge and skills to deliver the interventions as discussed in the training. Reframing the assumptions as hypotheses and articulating the theories underlying the causal chain produces a richer picture of the theory of change (Brown 2016), what measurement is needed (for example, providers' knowledge and skills at one or two times post-training) and how to use the experimental design of the impact evaluation to test the hypothesis (for example, comparing the performance of the providers trained with those in the control group).

We incorporate the explicit recognition and measurement of mediators and outcomes from implementation research into an archetypal theory of change for impact evaluation to suggest a framework for combining impact evaluation and implementation research. In this framework, the following elements are adopted from implementation research:

- Process evaluation of implementation moderators – facilitation, monitoring, feedback, adaptation, and relevance to providers and beneficiaries;
- A theory of change about the behavior of individuals and organizations implementing the intervention;
- Measurement and reporting of adherence to content and dose (fidelity); and
- Standardized reporting of the intervention, sufficient for replication.

Greater rigor in measuring implementation is adopted from impact evaluations using experimental or quasi-experimental study designs. Evaluation of both implementation and beneficiary outcomes can use experimental or quasi-experimental designs.

In reality, many impact evaluations incorporate some implementation research elements through a mixed-methods approach, providing qualitative information on the context and assessing why and how the results were obtained (positive, negative or null). However, they often miss opportunities that could provide a deeper understanding – more measures of fidelity, greater assessment of assumptions, more indicators along the causal chain. What defines implementation research is less about the methods and more about the question. Often, the question could be answered with an impact evaluation, but is not.

4. Case studies

We examined three published impact evaluation studies to understand how they did, or did not, incorporate implementation research approaches to measure and evaluate the delivery of the intervention under study. We selected two impact evaluations from an International Initiative for Impact Evaluation (3ie) program on the integration of HIV services (details in the appendix). These studies evaluate new interventions for integrating HIV treatment and care into PMTCT and maternal and neonatal child health services. We selected these studies because the implementers and researchers had more control over the interventions' implementation, and thus the scope for including implementation research questions, relative to the other studies in the program.

We selected a third case study from the published literature, in which researchers and implementers integrated implementation research and impact evaluation to build the case for an innovative approach to long-term care for people living with HIV.

For each case study, we examine how impact evaluation and implementation research questions and outcomes are combined, as well as missed opportunities.

4.1 Case study 1: evaluating the impact of community health worker integration into prevention of mother-to-child transmission of HIV services in Tanzania²

4.1.1 Purpose of the evaluation

Community health workers (CHWs) are frontline health paraprofessionals who can enhance the reach, coverage and quality of HIV services (Mwai et al. 2013); however, there is sparse evidence for leveraging this essential cadre to increase the effectiveness of PMTCT services. To test whether CHWs could enhance retention in PMTCT services, including HIV care and adherence to antiretroviral therapy (ART), the researchers conducted an impact evaluation of *Mama na Mtoto Pamoja* (Mother and Child Together) in Shinyanga, Tanzania, which integrated CHWs into PMTCT services.

4.1.2 Intervention

CHWs are part-time volunteers who provide referrals to care and information on health issues. In Shinyanga, CHWs work one to three days per week and receive remuneration for transport costs and completed reports. They each serve approximately 60 households in one community, visiting most households monthly, and are informally linked to a health facility in the area.

The implementation partner, Amref Health Africa, used four approaches to integration in the intervention sites:

- Formally linking CHWs to health facilities via weekly meetings;
- CHW-led ART adherence counseling for pregnant and postpartum women with HIV;
- CHW tracing of PMTCT clients lost to follow-up for HIV care; and
- Distribution of action birth cards, an interactive birth planning tool.

To implement the CHW intervention, the study team planned and executed comprehensive training on use of the action birth card, tracing women, adherence counseling and ethical issues related to confidentiality. The team provided additional financial compensation for these new activities, as well as compensation to CHWs in the control area (50 per cent of what those in the intervention area received) to mitigate compensation-related differentials in motivation. Amref supervised the CHWs in the intervention arm and offered refresher training when monitoring visits identified gaps in implementation. In addition, the study team held orientation meetings for health facility staff and the district health management team about the intervention and impact evaluation.

² This case study was developed from information published by Nance and colleagues (2017a; 2017b).

4.1.3 Evaluation design

The study used a cluster-randomized design to evaluate the short-term effectiveness of the CHW intervention to improve retention in care and ART adherence for HIV-infected pregnant and postpartum women. The researchers randomized 32 health facilities in Shinyanga, offering PMTCT services to the intervention group (n = 15) or comparison standard of care group (n = 17). Monitoring data were collected by the CHWs and their matched facilities. Intervention effectiveness was determined with a difference-in-differences analysis, based on clinical and pharmacy data from HIV-infected postpartum women at baseline and endline.

4.1.4 Findings and discussion

Program outputs included how many times the recommended topics were reported to have been discussed at CHW and healthcare worker meetings; the number of these meetings; the number of action birth cards distributed; and the number of defaulters identified, traced and returned to care. On the basis of these output measurements, the researchers concluded that the intervention “was smoothly implemented in both the treatment and control sites” (Nance et al. 2017a p.12).

Analysis of the expected impacts found that the intervention did not significantly change the primary outcome (retention in care at 90 days postpartum) or other outcomes (for example, ART initiation and timing during the pregnancy). However, the primary analyses and sensitivity analysis by intervention fidelity (at site level) both suggest that the intervention may have improved postpartum ART adherence, especially among women living in catchment areas where the intervention was implemented with higher intensity.

There was very little measurement of implementation. The study did not include measurement of knowledge and skills before or after the training or changes in the practice of desired behaviors by the CHWs. In addition, the CHWs had difficulty maintaining the paper logs used to track the number of visits and the topics discussed, and there were issues of low motivation related to competing projects that provided greater financial incentives to the CHWs.

4.1.5 Implementation research contribution

Our framework for the integration of impact evaluation and implementation research suggests that future iterations of the program should build in research questions and measures related to the fidelity, details and context of implementation. Potential research questions include the following:

- What training strategy for a cadre with low education levels results in improved knowledge and skills?
- What options can be tested to ensure supervision is sufficient and supportive? Do connection to a health facility and weekly meetings result in improved CHW performance?
- How acceptable are the various elements of the intervention to CHWs and facility-based providers?
- What is the quality of the services delivered by the CHWs?
- Was stigma prevented or minimized with the involvement of CHWs?

- How many CHWs delivered the full content and dose of the intervention, what are their characteristics and what was their exposure to implementation of the intervention (training, supervision, regular meetings)?
- What alternatives are there for collecting data on CHW outputs and performance?

In addition, the program could include an iterative or factorial design to test competing approaches for equipping and supporting CHWs to deliver the intervention with a high degree of fidelity.

4.2 Case study 2: improving ART adherence at reproductive and child health clinics integrating Option B+ in Tanzania³

4.2.1 Purpose of the evaluation

ART clinics in three East African countries found that implementing a minimally invasive, low-cost patient appointment and tracking system allowed them to promptly identify missing patients, facilitate the management of their workload, and promote sustainable and consistent clinic attendance by HIV-positive patients (Nyamusore et al. 2011; Mwatawala et al. 2012; Boruett et al. 2013). They also found that this correlated with medication adherence and clinical outcomes (Ross-Degnan et al. 2010).

To improve the retention of HIV-positive pregnant and postpartum women in PMTCT and HIV care in Tanzania, the Ministry of Health and Social Work distributed appointment books and a patient tracking system to all reproductive and child health clinics that were integrating HIV treatment services for HIV-positive pregnant and postpartum women as part of the rollout of Option B+. However, use of these tools was poorly understood and the system was not widely adopted.

To boost use of the appointment book and patient tracking system, the researchers undertook an impact evaluation of an implementation strategy for the introduction and rollout. They hypothesized that if there was appropriate orientation and supportive supervision during the first months after introduction, then providers' use of the intervention (appointment logs and patient tracking systems) would become routinized and the expected improvements in retention and adherence would be seen.

4.2.2 Intervention

The intervention consisted of training of trainers, clinic staff training and supportive supervision. Clinics in the control group received no intervention other than the Ministry of Health and Social Work's previous distribution of appointment books and patient-tracking registers.

4.2.3 Evaluation design

The researchers used matched-pair randomization, randomly assigning one district from each pair to the intervention and the other to standard care, in 24 reproductive and child health facilities in eight matched districts in the Mbeya region in Tanzania. The evaluation team collected quantitative data from pharmacy and clinic records, as well as qualitative data through interviews at baseline in intervention districts with clinic staff

³ This case study was developed from information published by Chalker (2017) and Ross-Degnan and colleagues (2017).

members, district staff members and women on ART at each clinic. At endline, they interviewed clinic staff members, district staff members, and women on ART at intervention and control facilities.

The researchers identified expected changes in patient behaviors regarding timely visits to the ART clinic and adherence to medication resulting from the changes in providers' practices. In their previous research, the selected behavioral outcomes had been shown to be linked to improved clinical outcomes and could be extracted from routine clinical data and aggregated to the clinic level.

In this impact evaluation, the intervention studied was not the appointment and patient tracking systems but the strategy for introducing it, which was intended "to empower the facility staff to use the appointment and patient-tracking tools" (Chalker 2017 p.4). The researchers measured the effectiveness of the implementation strategy in three ways. First, they assessed, during supervision visits, whether staff members understood and implemented the processes for using the appointment logs and tracking patients. Second, they used semi-structured interviews with staff at selected facilities to assess the extent to which facility staff monitored their own performance and adapted their practices.

Third, they used multiple statistical approaches to evaluate program effects on appointments and ART dispensing: interrupted time series with comparison series analysis, generalized estimating equation difference-in-difference estimation and Kaplan-Meier survival curves with accelerated Cox failure time models. Models were adjusted for possible pre- and post-intervention changes in outcomes in the control group, and all models were controlled for clustering.

4.2.4 Findings and discussion

Reports from supervision visits found that clinics' fidelity to the intervention differed and that the initial training on the appointment system was insufficient to change behavior. Many clinics received several supervision visits before they had successfully implemented the recommended changes, and not all recommended changes were successfully implemented.

Analysis of the quantitative data on appointments and ART dispensing found that the intervention significantly improved appointment-keeping and consistent availability of antiretroviral medicines for patients on long-term ART. However, there was considerable heterogeneity among facilities. Models estimating the likelihood of a missed visit in each facility found that in five of the intervention facilities and one control facility, the odds of a missed visit were significantly lower after the intervention. However, two intervention facilities and five control facilities demonstrated significantly higher odds of a missed visit in the post-intervention period.

Qualitative information suggested that regardless of the implementation strategy (training and supportive supervision or none), motivated staff, close liaisons with community organizations, and rapid identification and tracking of missing patients were components of success. Larger facilities in the urban area had less success, which the researchers suggest might be related to high staff turnover, a large volume of patients and fewer ties with community resources.

4.2.5 Implementation research contribution

In this case, the impact evaluation and intervention research were highly integrated. Rigorous impact evaluation methods (a combination of a comparison group and statistical methods) were used to measure the effectiveness of an implementation strategy (a training and supervision approach). The results from the impact evaluation design and statistical analysis point to the conclusion that positive changes in patient outcomes could be attributed to the intervention.

However, the implementation research, which took a detailed look at how the intervention was introduced and rolled out, uncovered factors not previously considered in the evaluation. For example, facility size and existing relationships between facilities and communities appeared to be potentially important predictors of patient outcomes. The additional findings from the implementation research are critical for tailoring and adapting the implementation strategy for specific settings.

4.3 Case study 3: piloting, evaluating and scaling up community ART support groups in Mozambique⁴

4.3.1 Purpose of the evaluation

The antiretroviral treatment program in Mozambique initially achieved considerable success enrolling HIV-positive patients into treatment. However, with patient attrition at 12 months estimated at 26 percent, there were serious concerns about the health system's ability to retain patients on treatment in the long term. To overcome patient-reported barriers to monthly drug refills for ART, the Tete Provincial Directorate of Health and Médecins Sans Frontières developed a novel, community-based, patient-centered ART model to pilot, evaluate, adapt and eventually scale up. Médecins Sans Frontières and the Ministry of Health designed the model to facilitate access to regular drug refills for people living with HIV in treatment and to reduce workload in the ART clinics.

4.3.2 Intervention

This patient-centered model used peer support groups called community ART groups (CAGs).⁵ People living with HIV formed groups based principally on belonging to the same social network and common needs as patients, with a maximum of six members per group. The CAG agreed to meet monthly to verify and register members' adherence to their treatment. Using pooled resources, each CAG sent one representative each month to the health facility to retrieve antiretroviral prescriptions for all group members. Every member was expected to serve as the group representative on a rotating basis, so each patient had contact with a health center at least once every six months. Group members could still visit the health center at any other time, for any reason.

At the facility, the CAG representative reviewed adherence and any other issues for each group member with a counselor or clinician, and a clinician prescribed ART and prophylactic drugs for each group member. The representative then returned to the community and distributed ART and other medicines to each patient. Healthcare workers, mainly counselors, monitored the group activities.

⁴ This case study was developed from information published by Decroo and colleagues (2011; 2013; 2017), Jobarteh and colleagues (2016), and Rasschaert and colleagues (2014a; 2104b).

⁵ The CAGs were later renamed "community adherence and support groups" (CASGs). We refer to them throughout this discussion as CAGs.

The first CAGs commenced in 2008 in rural health facility catchment areas where program managers had identified significant time and monetary costs associated with monthly clinic visits as an obstacle to adherence. In 2009, health officials in Tete extended the option to participate to patients living in semi-urban communities as a means of reducing time spent at health facilities and the stigma tied to frequent clinic visits. In 2011, high CAG and ART retention rates convinced the Ministry of Health and major partners to pilot the CAG strategy on a national scale (Decroo et al. 2013). Based on the results of this national pilot, the ministry expanded the program from a pilot to a national strategy (Jobarteh et al. 2016).

4.3.3 Evaluation design

The CAG model was based on patients' needs, which changed over time. Consequently, its implementation was a dynamic process, requiring continuous adaptations. To provide the data for adaptation, program managers and researchers used a mixed-methods approach to monitor delivery, uptake and effectiveness, combining implementation research and impact evaluation and analysis of qualitative and quantitative data.

Initial monitoring was based on three paper tools – the national patient-held appointment card, the clinic-based patient file and a group monitoring form. Information from the three monitoring tools was entered in an electronic database, and data were analyzed each month to track enrollment, patient follow-up and outcomes. Médecins Sans Frontières also conducted a qualitative study among the main stakeholders involved in the CAG model, focusing on the relevance, group dynamics and impact of the CAG model on individual patients, healthcare services and the broader community.

Initial monitoring and the qualitative study documented the fidelity of implementation, where and why gaps existed, and the factors that contributed to successful implementation, and tracked the expected outputs (patients retained in care). To evaluate impact, Jobarteh and colleagues (2016) analyzed routinely collected patient data on retention in care and mortality using a quasi-experimental design, employing a matched retrospective cohort. A propensity score-matched analysis was performed to assess differences in mortality and loss to follow-up between matched CAG and non-CAG members. Propensity scores were estimated using a random-effects logistic regression model.⁶

4.3.4 Findings and discussion

In the initial rollout between February 2008 and May 2010, of the 1,301 patients still in CAGs at the latter date, 1,269 (97.5%) remained in care, 30 (2%) died and 2 (0.2%) were lost to follow-up. The program managers concluded that early outcomes were highly satisfactory in terms of mortality and retention in care, lending support to such out-of-clinic approaches (Decroo et al. 2011). The qualitative study documented that the CAG model resulted in active patient involvement and empowerment, which resulted in improved ART retention. The study also noted that the CAG model prompted a

⁶ Covariates in the logistic regression model were selected a priori, based on their relationship to CAG eligibility and membership, and included sex, educational status, World Health Organization stage, year of ART initiation, age group, CAG eligibility, CD4 cell count category, weight (kg) and employment status. CAG eligibility was a matching criterion because some CAG-ineligible patients were found to have joined a CAG.

reorientation of healthcare services toward the community and strengthened community actions (Rasschaert et al. 2014b).

In the impact analysis, Jobarteh and colleagues (2016 p. 1221) found that non-CAG participants had higher loss to follow-up rates than matched CAG participants; however, there were no significant mortality differences between non-CAG and CAG participants. On this basis, the researchers made this conclusion:

These results support the Mozambican HIV program's decision to scale this model up to the entire country and demonstrate that a successful [CAG] program can be implemented on a large-scale by the [Ministry of Health] with support from implementing partners.

4.3.5 Implementation research contribution

Overall, the combination of implementation research and impact evaluation provide solid evidence for investing in and strengthening the CAG model. The impact evaluation, using a counterfactual, produced sufficiently robust evidence to support a recommendation to scale up the CAG model to improve the availability of long-term care to HIV patients and other patients with chronic illnesses. The implementation monitoring and research provided insights that led to program modifications and improvements, including permanent placement of counselors dedicated to the CAGs in most clinics. Medical eligibility criteria to join a CAG were used more flexibly, and patients were gradually more engaged in their healthcare, progressing to an active healthcare management collaboration in the clinics and community.

These findings resulted from more than one evaluation and were disseminated through multiple publications. So, while one study did not necessarily combine implementation research and impact evaluation, the program overall successfully combined the approaches and is an example of how initial planning could incorporate both from the beginning, even if they are conducted as separate research activities.

5. Conclusions

Integrating impact evaluation and implementation research requires incorporating and balancing questions and measures of effectiveness with questions and measures of the fidelity of implementation, assumptions about the causal pathway, facilitators and moderators of implementation and how they work together in the system being studied. Greater integration of impact evaluation and implementation research – and a liberal borrowing and exchange of methods – is a more robust and probably more efficient approach (though not a quicker one) for producing the desired information for evidence-informed action.

To avoid the pitfalls of post hoc and ad hoc approaches in implementation research, investigators can prospectively embed evaluation and measurement in the implementation process. Adopting some of the rigor applied in impact evaluation will result in more reliable and comprehensive evidence for decisions and program adaptations. Adding more assessments of implementation outcomes along the causal pathway and specifics of how, and how well, interventions are implemented will improve the usefulness and generalizability of impact evaluations.

6. Recommendations

Add research questions about the key elements of the implementation approach to impact evaluations. The inclusion of research questions regarding implementation in impact evaluations will stimulate and structure measurement of implementation outcomes, in addition to beneficiary outcomes. Some of the missed opportunities to measure implementation in the first case study were due to the limits set by the grant program, which determined the focus on beneficiary outcomes. In the second case study, the analysis of additional factors suggested by the embedded qualitative implementation research provided important insights into the successes and shortcomings of the implementation approach. In the third case study, an objective from the outset was to design a patient-focused intervention. Consequently, the evaluators built in extensive implementation research to understand how the intervention was delivered and received.

Include causal chain analysis about expected changes in professional behavior, organizational structures and relationships among stakeholders in impact evaluations to assess fidelity and appropriateness. The expected causal chain for achieving implementation was articulated in both 3ie studies, but there was insufficient measurement along the pathway or outcomes. Although some qualitative data provided information about fidelity and additional moderators of appointment and treatment adherence, a more iterative approach could have addressed some of the new factors. Stronger attention to process indicators along the causal pathway could have provided more rigorous evidence about fidelity of implementation. In part, this is a result of the lack of research questions about the implementation approach. In the second case study, the initial questions were primarily about implementation, and there were iterative research activities to understand changes in organizational structures and relationships among stakeholders. Once the model had been adapted on the basis of the implementation research, the impact evaluation was conducted.

Exploit experimental and quasi-experimental impact evaluation designs to conduct more rigorous implementation research. In the first case study, iterative or factorial designs could have been used to test competing approaches for equipping and supporting CHWs to deliver the intervention with a high degree of fidelity. The evaluation described in the second case study used an experimental design to assess the contribution of intervention approach (training and supervision). This allowed the evaluators to conclude, with confidence, that when the expected outcomes of the intervention approach were detected, change in professional behavior, organizational structures and relationships with stakeholders – and changes in practices of the delivery of services – could be attributed to the intervention approach. This conclusion would not have been possible without an experimental design.

Align resources and objectives. Implementation takes time, and research on implementation adds time and costs to an impact evaluation. Implementation of the CAG approach in Mozambique unfolded over four years, starting with formative research, piloting in rural areas, expansion in the rural areas and piloting in urban areas and, eventually, a national rollout. Ongoing implementation research used routine monitoring data, supervision visits and qualitative studies. After four years, the program conducted the impact evaluation. The integration of impact evaluation and implementation research

generally adds complexity and increases the human and financial resources required to produce the evidence.

Measure and report intervention costs. The cost of the intervention is important in assessing effectiveness, especially efficiency. Effectiveness studies sometimes provide evidence of cost-effectiveness (the cost per output or outcome achieved). This is useful to decision makers in choosing between interventions. When measures of effectiveness are not available to measure cost-effectiveness, the absolute cost of implementation is still highly relevant to implementers. Providing some kind of relative measure or point of comparison is very helpful, such as what similar interventions cost or the cost of another intervention that aims to affect the same outcome. A strength of the studies in the 3ie program is that they include the measurement of costs. (This information is available in the study reports.) The results show how expensive some of the tested interventions would be to replicate and scale.

Be more systematic about reporting on implementation research and impact evaluation studies. Without a complete, published description of interventions, programs cannot reliably implement interventions shown to be useful, and other researchers cannot replicate or build on research findings. To improve the completeness of reporting and the replicability of interventions, an international group of experts and stakeholders developed the template for intervention description and replication checklist and guide, known as TIDieR (Hoffman et al. 2014).

The standards for reporting implementation studies (StaRI) statement and checklist (Pinnock et al. 2017) further developed this approach for implementation research. The StaRI checklist is intended to improve reporting of implementation studies, by describing, on the one hand, the implementation approach and, on the other, the clinical, healthcare, global health or public health intervention being implemented. Overarching concerns when developing the StaRI checklist included balancing the need for detailed descriptions of interventions with publishing constraints, addressing the dual aims of reporting on the process of implementation and effectiveness of the intervention, and monitoring fidelity to an intervention while encouraging adaptation to suit diverse local contexts (Pinnock et al. 2015).

Appendix: 3ie thematic grant window for integration of HIV services

3ie is an international grant-making nongovernmental organization promoting evidence-informed decision-making for development policies and programs. 3ie's evidence program for the integration of HIV services aims to help bridge the knowledge gap of what works and why in HIV care and treatment, and specifically whether and how integration of HIV services with other health services could significantly improve the HIV and AIDS treatment cascade.⁷ In 2015, 3ie awarded five grants to fund projects that included the implementation and impact evaluation of pilots of under-researched HIV service integration interventions. The pilots aimed to improve linkage to care, adherence and/or retention. The goal of the associated impact evaluations was to contribute to a better understanding of what works, why, through what channels, and at what cost to maximize policy relevance and impact. Details of the five studies are shown below.

3ie integration of HIV services impact evaluations

Principal Investigator(s)	Organizations	3ie Report
1. Integration of community maternal, newborn and child health services with prevention of mother-to-child HIV transmission using community health workers to enhance retention in care and improve adherence to antiretroviral therapy in Tanzania		
Sandra McCoy	Amref Health Africa University of California, Berkeley	Evaluating the impact of community health worker integration into prevention of mother-to-child transmission of HIV services in Tanzania
2. Improving adherence to ART at maternal and child health clinics in Tanzania		
John Chalker, Dennis Ross-Degnan	Management Sciences for Health Harvard University	Improving ART adherence at reproductive and child health clinics integrating Option B+ in Tanzania
3. Integration of HIV Services in a Chronic Care Model in Côte d'Ivoire		
Stacie C. Stender	Jhpiego	Effectiveness evaluation of first-phase integrated chronic care model to improve prevention, care and support for people living with HIV in Côte d'Ivoire
4. Integration of expanded program for immunization and pediatric HIV services for improved coverage and patient outcomes in Zimbabwe		
Alexio Mangwiro, Brett Keller	Clinton Health Access Initiative	Integration of EPI and paediatric HIV services for improved ART initiation in Zimbabwe
5. Integration of HIV services into a community-based health programme in Tanzania		
Till Bärnighausen, Pascal Geldsetzer, Joel Francis	Harvard School of Public Health Management and Development for Health	Community delivery of antiretroviral drugs: a non-inferiority matched-pair pragmatic cluster-randomized trial in Dar es Salaam, Tanzania, Forthcoming

⁷ The HIV treatment cascade is a model that outlines the steps of care that people living with HIV go through, from initial diagnosis to achieving viral suppression (a very low level of HIV in the body), and shows the proportion of individuals living with HIV who are engaged at each stage.

References

Boruett, P, Kagai, D, Njogo, S, Nguhiu, P, Awuor, C, Gitau, L, Chalker, J, Ross-Degnan, D, Wahlstrom, R and Tomson, G, 2013. Facility-level intervention to improve attendance and adherence among patients on anti-retroviral treatment in Kenya: a quasi-experimental study using time series analysis. *BMC Health Services Research*, 13(242).

Brown, AN, 2016. Let's bring back theory to theory of change. 19 April 2016. *Evidence matters: improving development policy and practice: 3ie blog*. Available at: <http://blogs.3ieimpact.org/lets-bring-back-theory-to-theory-of-change/> [Accessed 1 May 2018].

Carroll, C, Patterson, M, Wood, S, Booth, A, Rick, J and Balain, S, 2007. A conceptual framework for implementation fidelity. *Implementation Science*, 2(40).

Chalker, J, 2017. *Improving ART adherence at reproductive and child health clinics integrating Option B+ in Tanzania*. 3ie Impact Evaluation Report 59. New Delhi: International Initiative for Impact Evaluation (3ie).

Curran, GM, Bauer, M, Mittman B, Pyne, JM and Stetler C, 2012. Effectiveness-implementation hybrid designs: combining elements of clinical effectiveness and implementation research to enhance public health impact. *Medical Care*, 50(3), pp.217–26.

Damschroder LJ, Aron DC, Keith RE, Kirsh SR, Alexander JA and Lowery JC, 2009. Fostering implementation of health services research findings into practice: a consolidated framework for advancing implementation science. *Implementation Science*, 4(1), p.50.

Decroo, T, Telfer, B, Biot, M, Maikéré, J, Dezembro, S, Cumba, LI, das Dores, C, Chu, K and Ford, N, 2011. Distribution of antiretroviral treatment through self-forming groups of patients in Tete Province, Mozambique. *Journal of Acquired Immune Deficiency Syndrome*, 56, pp.e39–e44.

Decroo, T, Lara, J, Rasschaert, F, Bermudez-Aza, EH, Couto, A, Candrinho, B, Biot, M, Maikere, J and Jobarteh, K, 2013. Scaling up community ART groups in Mozambique. *International STD Research & Reviews*, 1(2), pp.49–59.

Decroo, T, Telfer, B, Das Dores, C, White, RA, Dos Santos, N, Mkwamba, A, Dezembro, S, Joffrisse, M, Ellman, T and Metcalf, C, 2017. Effect of community ART groups on retention-in-care among patients on ART in Tete Province, Mozambique: a cohort study. *BMJ Open*, 7(8): e016800.

Easterly, W, 2009. The Civil War in Development Economics. 2 December 2009. *Aid Watch blog archive*. Available at: <http://www.nyudri.org/aidwatcharchive/2009/12/the-civil-war-in-development-economics/> [Accessed 8 August 2018].

Eccles, MP and Mittman, BS, 2006. Welcome to implementation science. *Implementation Science*, 1(1), p.1.

Fixsen, DL, Naoom, SF, Blase, KA, Friedman, RM and Wallace, F, 2005. *Implementation Research: A Synthesis of the Literature*. FMHI Publication #231. Tampa, FL: University of South Florida, Louis de la Parte Florida Mental Health Institute, The National Implementation Research Network.

Geng, EH, Peiris, D and Kruk, ME, 2017. Implementation science: Relevance in the real world without sacrificing rigor. *PLoS Medicine*, 14(4).

Gertler, PJ, Martinez, S, Premand, P, Rawlings, LB and Vermeersch, CMJ, 2011. *Impact Evaluation in Practice*. Washington DC: World Bank.

Habicht JP, Victora CG and Vaughan JP, 1999. Evaluation designs for adequacy, plausibility and probability of public health programme performance and impact. *International Journal of Epidemiology*, 28(1):pp.10–18

Hearn, S and Buffardi, AL, 2016. *What Is impact? A Methods Lab publication*. London: Overseas Development Institute.

Hemming, K, Haines, TP, Chilton, PJ, Girling, AJ and Lilford, RJ, 2015. The stepped wedge cluster randomised trial: rationale, design, analysis, and reporting, *BMJ*, 350.

Hoffmann, TC, Glasziou, PP, Boutron, I, Milne, R, Perera, R, Moher, D, Altman, DG, Barbour, V, Macdonald, M, Lamb, S, Dixon-Woods, M, McCulloch, P, Wyatt, J, Chan Phelan, A and Michie, S, 2014. Better reporting of interventions: template for intervention description and replication (TIDieR) checklist and guide. *BMJ*, p.348.

Jobarteh, K, Shiraishi, RW, Malimane, I, Samo Gudo, P, Decroo, T, Auld, AF, Macome, V and Couto, A, 2016. Community ART Support groups in Mozambique: the potential of patients as partners in care. *PLoS ONE*, 11(12).

Kalibala, S, Woelk, GB, Gloyd, S, Jani, N, Kay, L, Sarna, A, Okal, J, Ndwiga, C, Haberland, N and Sinai, I, 2016. Experiences in implementation and publication of operations research interventions: gaps and a way forward. *Journal of the International AIDS Society*, 19(5Suppl 4).

Mihalic, S, 2004. The importance of implementation fidelity. *Emotional & Behavioral Disorders in Youth*, 4, pp.83–86, 99–105.

Mwai, GW, Mburu, G, Torpey, K, Frost, P, Ford, N and Seeley, J, 2013. Role and outcomes of community health workers in HIV care in sub-Saharan Africa: a systematic review. *Journal of the International AIDS Society*, 16.

Mwatawala, S, Sando, D, Malele, R, Moshiro, C, Senyael, B, Somi, G, Maselle, A, Chalker, J and Ross-Degnan, D, 2012. Strengthening the appointment and tracking systems for patients on antiretroviral therapy in Tanzania: a strategy to optimize adherence to ART. *International Journal of Person Centered Medicine*, 2(4), pp.825–36.

Nance, N, McCoy, S, Ngilangwa, D, Masanja, J, Njau, P and Noronha, R, 2017a. *Evaluating the impact of community health worker integration into prevention of mother-to-child transmission of HIV services in Tanzania*, 3ie Impact Evaluation Report 61. New Delhi: International Initiative for Impact Evaluation (3ie).

- Nance, N, Pendo, P, Masanja, J, Ngilangwa, DP, Webb, K, Noronha, R and McCoy, SI, 2017b. Short-term effectiveness of a community health worker intervention for HIV-infected pregnant women in Tanzania to improve treatment adherence and retention in care: A cluster-randomized trial. *PLoS ONE*, 12(8).
- Nyamusore, J, Hinda, R, Mutabazi, V, Karema, C, Nsanzimana, S, Gaparayi, P, Chalker, J, Wagner, AK, Ross-Degnan, D, Ntaganira, J and Binagwaho, A, 2011. Effects of patient tracking systems and providers incentives on patient appointment keeping in Rwanda. Presented at the Third International Conference for Improving Use of Medicine, Antalya, Turkey, November 2011. Available at: <<https://slideplayer.com/slide/6853509/>> [Accessed 14 September 2018].
- Odeny, TA, Padian, N, Doherty, MG, Baral, S, Beyrere, C, Ford, N and Gene, E, 2015. Definitions of implementation science in HIV/AIDS. *The Lancet HIV*, 2(5), pp.e178–e180.
- Padian, NS, Holmes, CB, McCoy, SI, Lyerla, R, Bouey, PD and Goosby, EP, 2011. Implementation science for the US President’s Emergency Plan for AIDS Relief (PEPFAR). *Journal of Acquired Immune Deficiency Syndromes*, 56(3), pp.199–203.
- Parry, G and Power, M, 2016. To RCT or not to RCT? The ongoing saga of randomised trials in quality improvement. *BMJ Quality & Safety*, 25, pp.221–23.
- Patsopoulos, NA, 2011. A pragmatic view on pragmatic trials. *Dialogues in Clinical Neuroscience*, 13(2), pp.217–24.
- Peters, DH, Adam, T, Alonge, O, Akua Agyepong, I and Tran, N, 2013. Implementation research: what it is and how to do it. *BMJ*, 347.
- Pinnock, H, Barwick, M, Carpenter, CR, Eldridge, S, Grandes, G, Griffiths, CJ, Rycroft-Malone, J, Meissner, P, Murray, E, Patel, A and Sheikh, A, 2017. Standards for Reporting Implementation Studies (StaRI) Statement. *BMJ*, 356.
- Pinnock, H, Epiphaniou, E, Sheikh, A, Griffiths, C, Eldridge, S, Craig, P and Taylor, S, 2015. Developing standards for reporting implementation studies of complex interventions (StaRI): a systematic review and e-Delphi. *Implementation Science*, 10, p.42.
- Proctor, E, Silmere H, Raghavan R, Hovmand P, Aarons G, Bunger A, Griffey, R and Hensley, M, 2011. Outcomes for implementation research: conceptual distinction, measurement challenges, and research agenda. *Administration and Policy in Mental Health*, 38, pp.65–76.
- Rabin, BA, Brownson, RC, Haire-Joshu, D, Kreuter, MW and Weaver, NL, 2008. A glossary for dissemination and implementation research in health. *Journal of Public Health Management Practice*, 14(2), pp.117–23.
- Rasschaert, F, Decroo, T, Remartinez, D, Telfer, B, Lessitala, F, Biot, M, Candrinho, B and Van Damme, W, 2014a. Adapting a community-based ART delivery model to the patients’ needs: a mixed methods research in Tete, Mozambique. *BMC Public Health*, 14, p.364.

Rasschaert F, Telfer B, Lessitala F, Decroo T, Remartinez D, Biot, M, Candrinho, B, Mbofana, F and Van Damme, W, 2014b. A qualitative assessment of a community antiretroviral therapy group model in Tete, Mozambique. *PLoS ONE*, 9(3).

Ridde V, 2016. Need for more and better implementation science in global health. *BMJ Global Health*, 1.

Ross-Degnan, D, Chalker, J, Liana, J, Kajoka, MD, Valimba, R, Kimatta, S, Dillip, A, Vialle-Valentin, C, Embrey, M, Lieber, R and Johnson, K, 2017. A group randomized trial using an appointment system to improve adherence to ART at reproductive and child health clinics implementing Option B+ in Tanzania. *PLoS ONE*, 12(9).

Ross-Degnan, D, Pierre-Jacques, M, Zhang, F, Tadege, H, Gitau, L, Ntaganira, J, Balikuddembe, R, Chalker, J and Wagner, AK, 2010. Measuring adherence to antiretroviral treatment in resource-poor settings: the clinical validity of key indicators. *BMC Health Services Research*, 10, p.42.

Singal, AG, Higgins, P and Waljee AK, 2014. A primer on effectiveness and efficacy trials. *Clinical and Translational Gastroenterology*, 5(1), p.e45.

White, H, 2009. *Some reflections on current debates in impact evaluation*. 3ie Working Paper 1. New Delhi: International Initiative for Impact Evaluation (3ie).

Other publications in the 3ie working paper series

The following papers are available from

<http://www.3ieimpact.org/en/publications/workingpapers/>

Synthesis of impact evaluations of the World Food Programme's nutrition interventions in humanitarian settings in the Sahel, 3ie Working Paper 31. Kaul, T, Husain, S, Tyrell, T and Gaarder, M, 2018.

Community-driven development: does it build social cohesion or infrastructure? A mixed-method evidence synthesis, 3ie Working Paper 30 White, H, Menon, R and Waddington, H, 2018.

Evaluating advocacy: an exploration of evidence and tools to understand what works and why. 3ie Working Paper 29. Naeve, K, Fischer-Mackey, J, Puri, J, Bhatia, R and Yegbemey, R, 2017.

3ie evidence gap maps: a starting point for strategic evidence production and use, 3ie Working Paper 28. Snilstveit, B, Bhatia, R, Rankin, K and Leach, B (2017)

Examining the evidence on the effectiveness of India's rural employment guarantee act, 3ie Working Paper 27. Bhatia, R, Chinoy, SL, Kaushish, B, Puri, J, Chahar, VS and Waddington, H (2016)

Power calculation for causal inference in social science: sample size and minimum detectable effect determination, 3ie Working Paper 26. Djimeu, EW and Houndolo, DG (2016)

Evaluations with impact: decision-focused impact evaluation as a practical policymaking tool, 3ie Working Paper 25. Shah, NB, Wang, P, Fraker, A and Gastfriend, D (2015)

Impact evaluation and policy decisions: where are we? A Latin American think-tank perspective, 3ie Working Paper 24. Baanante, MJ and Valdivia, LA (2015)

What methods may be used in impact evaluations of humanitarian assistance? 3ie Working Paper 22. Puri, J, Aladysheva, A, Iversen, V, Ghorpade, Y and Brück, T (2014)

Impact evaluation of development programmes: experiences from Viet Nam, 3ie Working Paper 21. Nguyen Viet Cuong (2014)

Quality education for all children? What works in education in developing countries, 3ie Working Paper 20. Krishnaratne, S, White, H and Carpenter, E (2013)

Promoting commitment to evaluate, 3ie Working Paper 19. Székely, M (2013)

Building on what works: commitment to evaluation (c2e) indicator, 3ie Working Paper 18. Levine, CJ and Chapoy, C (2013)

From impact evaluations to paradigm shift: A case study of the Buenos Aires Ciudadanía Porteña conditional cash transfer programme, 3ie Working Paper 17. Agosto, G, Nuñez, E, Citarroni, H, Briasco, I and Garcette, N (2013)

Validating one of the world's largest conditional cash transfer programmes: A case study on how an impact evaluation of Brazil's Bolsa Família Programme helped silence its critics and improve policy, 3ie Working Paper 16. Langou, GD and Forteza, P (2012)

Addressing attribution of cause and effect in small n impact evaluations: towards an integrated framework, 3ie Working Paper 15. White, H and Phillips, D (2012)

Behind the scenes: managing and conducting large scale impact evaluations in Colombia, 3ie Working Paper 14. Briceño, B, Cuesta, L and Attanasio, O (2011)

Can we obtain the required rigour without randomisation? 3ie Working Paper 13. Hughes, K and Hutchings, C (2011)

Sound expectations: from impact evaluations to policy change, 3ie Working Paper 12. Weyrauch, V and Langou, GD (2011)

A can of worms? Implications of rigorous impact evaluations for development agencies, 3ie Working Paper 11. Roetman, E (2011)

Conducting influential impact evaluations in China: the experience of the Rural Education Action Project, 3ie Working Paper 10. Boswell, M, Rozelle, S, Zhang, L, Liu, C, Luo, R and Shi, Y (2011)

An introduction to the use of randomised control trials to evaluate development interventions, 3ie Working Paper 9. White, H (2011)

Institutionalisation of government evaluation: balancing trade-offs, 3ie Working Paper 8. Gaarder, M and Briceño, B (2010)

Impact evaluation and interventions to address climate change: a scoping study, 3ie Working Paper 7. Snilstveit, B and Prowse, M (2010)

A checklist for the reporting of randomised control trials of social and economic policy interventions in developing countries, 3ie Working Paper 6. Bose, R (2010)

Impact evaluation in the post-disaster setting, 3ie Working Paper 5. Buttenheim, A (2009)

Designing impact evaluations: different perspectives, contributions, 3ie Working Paper 4. Chambers, R, Karlan, D, Ravallion, M and Rogers, P (2009) [Also available in Spanish, French and Chinese]

Theory-based impact evaluation, 3ie Working Paper 3. White, H (2009) [Also available in French and Chinese]

Better evidence for a better world, 3ie Working Paper 2. Lipsey, MW (ed.) and Noonan, E (2009)

Some reflections on current debates in impact evaluation, 3ie Working Paper 1. White, H (2009)